STAT 300 Chapter 4

Zepu Zhang
February 27, 2012

1 Probability density function (pdf) of a continuous distribution

In contrast to discrete distributions, which is specified by the "probability mass" $p(x)$ for each possible value of X, continuous distributions can not be specified this way. The reasons? Informally speaking, there are too many possible values for X. Actually, $P(X = x) = 0$, for any particular value x.

Solution: specify the probability of an interval, $P(a < X < b)$.

The density function, $f(x)$ is such that

1. $f(x) \geq 0$.
2. $P(a \leq X \leq b) = \int_a^b f(x)\,\mathrm{d}x$.
3. $\int_{-\infty}^{\infty} f(x)\,\mathrm{d}x = 1$.

   Note the logic here: knowing a distribution completely is equivalent to knowing $P(a \leq X \leq b)$ for any $a < b$. We find a function $f(x)$ that conveys this knowledge.

   Suppose on some interval $[a, b]$, the "chance" of X falling anywhere is the same. Then $P([a, b]) = \int_a^b f(x)\,\mathrm{d}x = (b - a)f(x)$. It follows that $f(x) = \frac{P([a,b])}{b-a}$. Hence $f(x)$ indicates the "density" of probability on the interval $[a, b]$. Although we can't expect the "chance" to be the same on all intervals, a similar argument still makes sense if we consider increasingly small (narrow) intervals. This explains why the function $f(x)$ indicates the density.

   We often use $p(x)$ for pmf and $f(x)$ for pmf. But this distinction is not absolute.

Density curve is the plot for the function $f(x)$, and has the same properties as above. Visually speaking,

1. No part is below the horizontal axis.
2. $P(a \le X \le b)$ equals the area on the interval $[a, b]$ below the curve.
3. The total area below the curve is 1.

An important technical point: for a continuous rv (meaning random variable) X, the probability that X takes any particular value is 0.

$$P(X = a) = \int_a^a f(x)\,\mathrm{d}x = 0$$

In a continuous distribution, no single value has a positive probability. Only intervals have positive probabilities.

Implication: when we write the probability (of a continuous rv) for an interval, openness or closeness at the end points does not matter.

$$P\big([a, b]\big) = P\big((a, b]\big) = P\big((a, b)\big) = P\big([a, b)\big)$$

For a continuous distribution, we talk about the probability of an interval, not that of any single value.

Example The simplest continuous distribution is the uniform distribution: X takes values on interval $[a, b]$ and is equally likely anywhere on this interval.

$$f(x; a, b) = \begin{cases} \frac{1}{b-a}, & a \le x \le b \\ 0 & \text{otherwise} \end{cases}$$

What does the density curve of a uniform distribution look like?

2 Cumulative distribution function (cdf)

Definition (for both discrete and continuous rv's):

$$F(x) = P(X \le x)$$

2.1 Obtain cdf from pdf

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(y)\,\mathrm{d}y$$

On the pdf plot, $F(x)$ is the area under the density curve to the left of x. Note it's a "$\leq$", not "$<$", in the definition of cdf. (This detail is especially important for discrete distributions.)

The plot of $F(x)$ is a non-decreasing curve.

Example The uniform distribution.

Example Ex. 4.7 in Chap. 4.2.

Example Ex. 4.9 in Chap. 4.2.

Note Whenever possible, use a graph to help yourself.

2.2 Obtain pdf from cdf

Suppose the continuous X has pdf $f(x)$ and cdf $F(x)$. At any x where the derivative $F'(x)$ exists, we have (or define)

$$f(x) = F'(x)$$

This is going in the opposite direction of the definition of $F(x)$. Note all the conditions:

1. This is for continuous variable.
2. The pdf exists. (The cdf always exists; pdf may not. The pdf requires $F'(x)$ to exist.)

The continuous distributions we encounter will all meet these requirements.

For our purpose, the caution is mainly about discrete distributions. Discrete distributions do not have pdf (their counterpart is pmf); the cdf of a discrete distribution is not everywhere differentiable.

Example Uniform distribution.

Example The distribution whose cdf is $F(x) = 1 - e^{-3x}$ for $x \geq 0$ and $F(x) = 0$ for $x < 0$.

Note Discrete rv has probability mass; continuous rv has probability density. Use summation and subtraction for discrete rv; use integral and derivative for continuous rv.

2.3 Obtain probabilities from cdf

$$P(X > a) = 1 - F(a)$$
$$P(a < X \le b) = P(X \le b) - P(X \le a) = F(b) - F(a)$$

2.4 Percentiles

The $(100p)$th percentile, $0 < p < 1$, is the value x such that $F(x) = p$. For example, the 74th percentile is the value x such that $F(x) = 0.74$. This is also called the p quantile.

Actually it's more convenient to use "quantile", because you'll announce p directly without converting to $100p$.

On a pdf curve, find a quantile by requiring the area to the left of x to be p.

On a cdf curve, the p quantile is the x whose y is p, that is, $F^{-1}(p)$ (the quantile function is the inverse of the cdf). It's more direct.

Example Find percentiles on a cdf curve.

Note Percentiles and quantiles for a discrete rv are defined similarly, but their determination is trickier due to the discreteness.

3 Mean and variance

3.1 The expected value of a function of X

$$E\big[h(X)\big] = \int_{-\infty}^{\infty} h(x) \cdot f(x)\,\mathrm{d}x \tag{1}$$

Exercise Compare with the "weighted average" definition in the discrete case and observe the analogy.

3.2 Mean

Its definition is obtained by taking $h(X)$ to be X in (1), that is,

$$E[X] = \int_{-\infty}^{\infty} x \cdot f(x)\,\mathrm{d}x$$

Example Uniform distribution.

Example Ex. 4.10 in Chap. 4.2.

3.3 Variance

$$\mathrm{var}(X) \equiv E\big[(X-\mu_X)^2\big] = \int_{-\infty}^{\infty} (x-\mu)^2 \cdot f(x)\,\mathrm{d}x \quad (2)$$

It is almost always preferred to use the formula

$$\mathrm{var}(X) = E\big[X^2\big] - \mu^2$$

4 Uniform distribution

(This is the simplest continuous distribution.)

PDF:

$$f(x) = \begin{cases} \frac{1}{b-a}, & a \le x \le b \\ 0, & \text{otherwise} \end{cases}$$

where $-\infty < a < b < \infty$.

Example Derive the cdf, mean, and variance of the uniform distribution. (You should be familiar with these results.)

5 Normal distribution

This is the most important distribution in all of probability and statistics. Its importance stems from very high visibility: many (continuous) numerical populations have approx a normal distribution. Even discrete numerical populations often have a shape similar to normal. Measurement errors or things of a similar nature (random fluctuation influenced by many factors) usually have a normal distribution. There are theoretical reasons for this high visibility: Central Limit Theorem (later).

Definition A continuous rv X defined on $(-\infty, \infty)$ has a normal distribution with parameters μ and σ^2 if its pdf is

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

This distribution is almost always written as $N(\mu, \sigma^2)$. That X has this distribution is usually written as $X \sim N(\mu, \sigma^2)$.

μ and σ^2 are the mean and variance of this distribution.

Exercise Verify $E(X) = \mu$ and $\text{var}(X) = \sigma^2$, using the definitions of mean and variance and the formula of $f(x)$ above.

The density curve is a nice "bell shape", centered at and symmetric about μ. σ is a signature length indicating the "spread" of the distribution.

Note You should remember the normal pdf formula.

5.1 Standard normal distribution

5.1.1 pdf

A standard normal variable is often denoted by Z: $Z \sim N(0, 1)$.

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}$$

5.1.2 cdf

Often denoted by $\Phi(z)$:

$$\Phi(z) = P(Z \leq z)$$

There is no analytical form. Calculation of it (numerically) is a basic function of every statistical software package. Also pre-computed and tabulated.

5.1.3 Learn to use a normal cdf table

Table A.3.

Example Ex. 4.13 in Chap. 4.3. Calculating probabilities of intervals.

Example Ex. 4.14 in Chap. 4.3. Finding percentiles.

5.1.4 The z_α notation

z_α is the value such that $P(Z > z_\alpha) = \alpha$. In other words, it is the $1 - \alpha$ quantile.

> z_α is very useful in hypothesis tests (later). In that context it is called the "critical value".

Note Always visualize.

5.2 Nonstandard normal distribution

5.2.1 Conversions between nonstandard and standard normal distributions

Let $X \sim N(\mu, \sigma^2)$ and $Z \sim N(0, 1)$.

Standardizing X leads to Z:

$$X \sim N(\mu, \sigma^2) \quad \Rightarrow \quad Z = \frac{X - \mu}{\sigma} \sim N(0, 1)$$

Scaling and shifting Z leads to X:

$$Z \sim N(0, 1) \quad \Rightarrow \quad X = \mu + \sigma Z \sim N(\mu, \sigma^2)$$

(First, scale Z by σ, then shift by μ.)

The above suggests that a linear transform of Z is a (nonstandard) normal variable, say X. A linear transform of X is again a linear transform of Z. Therefore we have the following general fact:

A linear function of a normal variable is a normal variable. (Later we'll learn a more general fact, which says that a linear combination of normal random variables is a normal random variable.)

All computations about a nonstandard normal distribution is relegated to the standard normal via standardization.

Convert probability statements of X to those of Z (page

149, box, 7th ed.; page 157, box, 8th ed):

$$\begin{aligned} P(a \le X \le b) &= P\Big(\frac{a-\mu}{\sigma} \le Z \le \frac{b-\mu}{\sigma}\Big) \\ &= \Phi\Big(\frac{b-\mu}{\sigma}\Big) - \Phi\Big(\frac{a-\mu}{\sigma}\Big) \\ P(X \le a) &= \Phi\Big(\frac{a-\mu}{\sigma}\Big) \\ P(X > b) &= 1 - \Phi\Big(\frac{b-\mu}{\sigma}\Big) \end{aligned}$$

Note Be very familiar with these conversions.

Example Ex. 4.16 in Chap. 4.3.

5.2.2 Compute quantiles

It is by the same strategy that we find percentiles (or quantiles) of a nonstandard normal.

Note quantile function is the inverse of CDF, e.g. the 0.78 quantile is $F^{-1}(0.78)$.

Denote $F_X^{-1}(q)$, $0 < q < 1$, by q_X. By definition,

$$F_X(q_X) = q$$

that is

$$P(X \le q_X) = q,$$

or equivalently,

$$P\left(Z = \frac{X-\mu}{\sigma} \le \frac{q_X - \mu}{\sigma}\right) = q$$

hence

$$\frac{q_X - \mu}{\sigma} = \Phi^{-1}(q)$$

and

$$q_X = \mu + \sigma\Phi^{-1}(q)$$

or $q_X = \mu + \sigma z_{1-q}$.

Percentiles of a nonstandard normal:

$$\begin{aligned} &(100p)\text{th percentile of } N(\mu,\sigma) \\ &\qquad = \mu + \big[(100p)\text{th percentile of } N(0,1)\big] \cdot \sigma \end{aligned}$$

Empirical rule: p. 151 (7th ed) or p. 159 (8th ed). (Good to know; no need to memorize.)

Example Ex. 4.18 in Chap. 4.3.

Note This happens a lot in applications: a quantity that is positive by nature is modeled by a normal distribution. In theory, this can only be approximate. In practice, it's often good enough if the normal distribution with the given mean and variance has virtually zero area to the left of value 0.

Example Ex. 4.17 in Chap. 4.3.

6 The exponential distribution

A very important distribution for positive variables.

pdf: with parameter $\lambda > 0$,

$$f(x) = \lambda e^{-\lambda x}, \quad x \geq 0$$

cdf:

$$F(x) = 1 - e^{-\lambda x}, \quad x \geq 0$$

The first two moments:

$$E(X) = \frac{1}{\lambda}, \quad \mathrm{var}(X) = \frac{1}{\lambda^2}$$

Plot the pdf curve.

Note Some books parameterize the pdf as $\frac{1}{\lambda}e^{-x/\lambda}$. Then $E(X)$, $\mathrm{var}(X)$, and interpretation of λ change accordingly.

Exercise (1) Derive the exponential cdf from its pdf, and vice versa.
(2) Derive $E(X)$ and $\mathrm{var}(X)$. (Calculus needed.)

7 Useful `R` functions

`runif`

`rnorm, dnorm, pnorm, qnorm`